

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a postprint version which may differ from the publisher's version.

For additional information about this publication click this link.

<https://hdl.handle.net/2066/231143>

Please be advised that this information was generated on 2021-06-14 and may be subject to change.

## Review

## RNA-Centric Methods: Toward the Interactome of Specific RNA Transcripts

Cathrin Gräwe,<sup>1,2</sup> Suzan Stelloo,<sup>1,2</sup> Femke A.H. van Hout,<sup>1</sup> and Michiel Vermeulen<sup>1,\*,@</sup>

**RNA–protein interactions play an important role in numerous cellular processes in health and disease. In recent years, the global RNA-bound proteome has been extensively studied, uncovering many previously unknown RNA-binding proteins. However, little is known about which particular proteins bind to which specific RNA transcript. In this review, we provide an overview of methods to identify RNA–protein interactions, with a particular focus on strategies that provide insights into the interactome of specific RNA transcripts. Finally, we discuss challenges and future directions, including the potential of CRISPR–RNA targeting systems to investigate endogenous RNA–protein interactions.**

**Importance of Studying RNA–Protein Complexes**

The various types of RNA, such as mRNAs, rRNAs, and long noncoding (lnc)RNAs, are involved in a multitude of cellular processes. The classical view of RNAs serving merely as a template for protein synthesis has been expanded to catalytic, structural, and regulatory functions [1,2]. RNAs interact with **RNA-binding proteins (RBPs)** (see [Glossary](#)), which play an essential role in RNA fate and function, such as in mRNA processing and translation [3]. In addition, RNA binding can also change the fate and function of proteins, for example by regulating protein stability and localization [4,5]. The importance of these interactions also becomes clear in the context of diseases that are characterized by perturbed RNA–protein interactions. Any alteration in expression level, structure, localization, or modification of either RNA or protein may disrupt these interactions and result in deregulation of the involved cellular processes [6,7]. For example, the RNA modification *N*6-methyladenosine (m6A) affects protein binding, and alterations in m6A levels are frequently observed in cancer [8–10].

To characterize RNA–protein interactions, a rapidly expanding toolbox of methods is available, often classified into protein-centric and RNA-centric methods. Protein-centric methods, such as crosslinking immunoprecipitation (CLIP), rely on immunoprecipitation of the protein of interest followed by sequencing of the associated RNAs [11,12]. Conversely, RNA-centric methods rely on isolation of proteins associated with the RNA of interest. These methods led to the identification of many RBPs, including unexpected proteins, such as metabolic enzymes and transcription factors (reviewed in [5]). In this review, we provide an overview of the current and emerging RNA-centric approaches ([Figure 1](#)). In addition, we highlight the advantages, disadvantages, and challenges that have to be overcome to improve current methods. Furthermore, we discuss potential technical innovations to provide robust insights into endogenous protein–RNA interactomes.

**Global RNA-Binding Proteome**

To gain insight into the function of RNA–protein interactions, several methods have been developed in recent years to identify the global RNA-bound proteome. These methods preserve cellular RNA–protein interactions by **crosslinking**, which is then followed by cell lysis and purification of the RNA–protein complexes. Purification of all cellular RNA–protein complexes can be done in multiple ways; for example, total RNA-associated protein purification (TRAPP) is based on

**Highlights**

RNA–protein interactions play a role in various cellular processes, and methods to study these interactions are expanding.

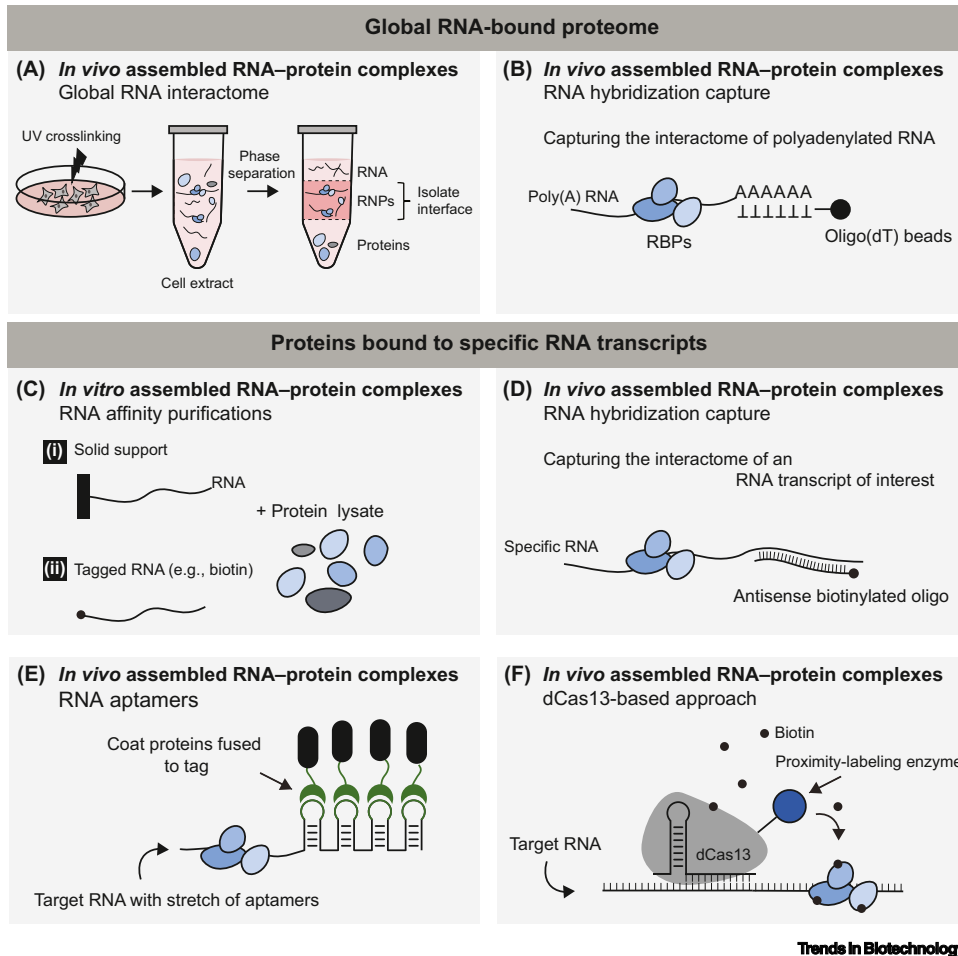
RNA–protein complexes can be isolated, after which bound proteins can be identified by western blot or mass spectrometry.

Despite the variety of existing methods, it remains challenging to characterize the global interactome of a specific RNA transcript in their native context and upon cellular perturbations.

Promising technologies use RNA-targeting systems coupled with proximity-dependent biotinylation to identify RNA–protein interactions *in vivo*.

<sup>1</sup>Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Oncode Institute, Radboud University Nijmegen, 6525 GA Nijmegen, The Netherlands  
<sup>2</sup>Co-first authors

\*Correspondence: [michiel.vermeulen@science.ru.nl](mailto:michiel.vermeulen@science.ru.nl) (M. Vermeulen).  
@Twitter: @LabVermeulen (M. Vermeulen).



Glossary

**Aptamers:** short, single-stranded nucleic acids that bind specifically to their target proteins, often with high affinity and specificity.

**Binding affinity:** the affinity of an interaction indicates how strong a given interaction is. A common measure for this is the dissociation constant ( $K_d$ ). The lower the  $K_d$  value, the stronger the interaction.

**Catalytically dead Cas13 (dCas13):** developed by mutating the two nuclease domains of Cas13. dCas13 retains affinity for RNA, but lacks RNA-cleavage activity.

**Coat proteins:** small proteins that specifically interact with RNA stem loop structures.

**Crosslinking:** formaldehyde or UV light can be used to preserve RNA-protein interactions. Formaldehyde forms crosslinks between proteins and RNA as well as between proteins, whereas UV forms crosslinks exclusively between proteins and RNA.

**Proximity-labeling enzyme:** an enzyme that facilitates the attachment of biotin to proximal proteins. Proximity-labeling enzymes can be categorized in biotin ligases [BioID, (mini)TurboID, and BASU] and peroxidases (APEX and APEX2).

**RNA-binding proteins (RBPs):** proteins that bind RNA, thereby forming regulatory complexes that control numerous cellular processes.

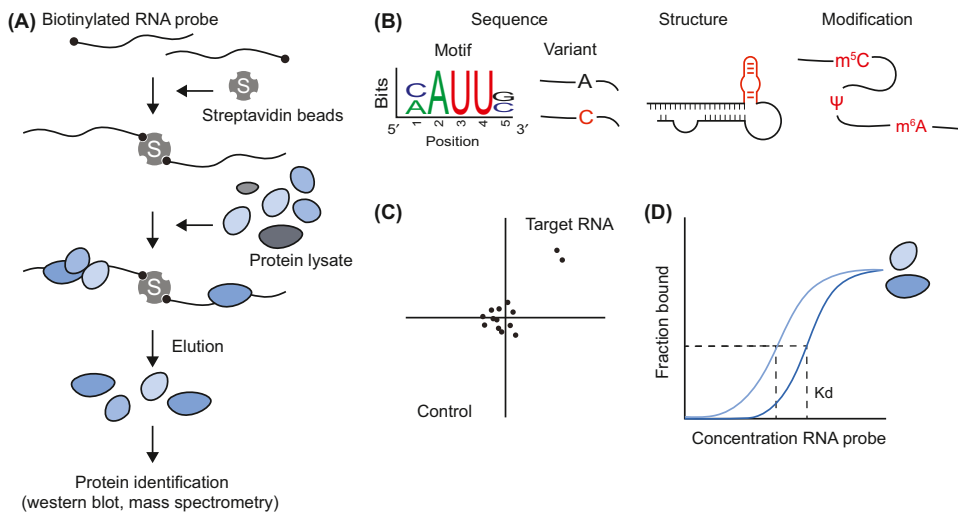
Figure 1. Schematic Illustration of RNA-Centric Methods for Detecting RNA-Protein Interactions. (A) The global RNA-bound proteome can be isolated by organic phase separation. (B) RNA hybridization capture methods to study proteins bound to polyadenylated RNA use oligo-dT coupled to beads. (C) For RNA affinity purifications, an RNA bait is bound to a solid support (i) or immobilized on beads (ii). The protein extract is then added to capture *in vitro* assembled RNA-protein interactions. (D) RNA hybridization capture methods use antisense DNA oligonucleotides that anneal to a specific RNA transcript. (E) RNA aptamer strategies rely on the co-expression of an aptamer-tagged RNA and coat proteins fused to a tag in living cells. The coat proteins can either be recognized by specific antibodies allowing to pulldown the target RNA or can be fused to proximity biotinylation enzymes, which label neighboring proteins. (F) In dCas13-based approaches, a guide RNA (gRNA) recruits catalytically inactive Cas13 (dCas13) fused to a biotin-labeling enzyme specifically to an RNA of interest. Upon addition of biotin, the biotin-labeling enzyme biotinylates proteins in its proximity. Abbreviations: RBP, RNA-binding proteins; RNPS, ribonucleoprotein particles.

silica-based matrices that bind nucleic acids [13]. Other methods, such as orthogonal organic phase separation (OOPS) [14], protein-crosslinked RNA extraction (XRNAX) [15], or phenol toluol extraction (Ptex) [16], rely on extraction of crosslinked RNA-protein complexes by organic phase separation (Figure 1A). Alternatively, polyadenylated RNA-protein complexes can be purified using oligo-dT coupled beads, a method often referred to as RNA interactome capture (RIC) (Figure 1B) [17,18]. This approach has been applied to study various biological processes. For instance, Liepelt and colleagues studied changes of the RNA-bound proteome upon lipopolysaccharide (LPS) stimulation in macrophages, thereby gaining important insights into signaling pathways involved in innate immune responses [19]. In another study, oligo-dT capture was applied to determine changes in the RNA-bound proteome after inhibition of RNA demethylases, thereby

identifying m6A-regulated RBPs [20]. Furthermore, oligo-dT capture approaches have been adapted to map the exact RNA-binding sites of proteins [21]. These and similar studies allowed the characterization of the global RNA-bound proteome, identifying known as well as previously unknown, unexpected RBPs [5]. However, some questions remain that cannot be addressed by studying the global RNA-bound proteome. What is the function of RNA–protein interactions and how are they regulated? To answer these questions, interactomes for specific RNA transcripts need to be characterized, thus providing new insights into the function of RNAs, proteins, and their interactions.

### *In Vitro* Assembled RNA–Protein Complexes

Numerous *in vitro* RNA affinity purifications approaches have been developed to identify proteins interacting with RNA sequences of interest (Figure 1C). Generally, an RNA bait is immobilized on beads and incubated with proteins extracted from tissues or cells. Then, unbound proteins are washed away and bound proteins are purified and identified by western blot or mass spectrometry (Figure 2A) [22]. This strategy has been widely adapted and different approaches have been developed. For example, immobilization on beads can be achieved using either chemical modifications or RNA tags; RNA baits can either be synthesized or *in vitro* transcribed; and different kinds of protein mixture can be used, such as nuclear or cytoplasmic extracts (reviewed in [22]). These approaches demonstrate that RNA affinity purifications can be applied to various biological questions. After the first description by Grabowski and Sharp in 1986, who studied proteins binding to small nucleolar RNAs [23], RNA affinity purifications have been used to identify



#### Trends in Biotechnology

**Figure 2. RNA Affinity Purifications.** (A) Schematic workflow of RNA affinity purifications. In this example, a biotin-tagged RNA bait is immobilized on streptavidin beads. Alternatively, the RNA bait is immobilized on a solid support as shown in Figure 1A in the main text. Protein extract is incubated with the RNA bait to capture *in vitro* assembled RNA–protein interactions. Subsequently, unbound proteins are removed by stringent washes and bound proteins are eluted. The eluted proteins are further processed for western blot or mass spectrometry analysis. (B) The binding of proteins to RNA can be influenced by the RNA sequence (e.g., motif or nucleotide variant), its secondary structure, and RNA modification, including 5-methylcytosine (m<sup>5</sup>C), pseudouridine (Ψ), and N<sup>6</sup>-methyladenosine (m<sup>6</sup>A), and can be studied by using RNA affinity purifications. (C) A graphical illustration of a scatter plot showing protein ratios obtained in quantitative mass spectrometry analysis. In semiquantitative mass spectrometry approaches, proteins or peptides are isotopically labeled to differentiate the experimental and control sample. The protein ratio of the two labels (experimental sample over control sample) is calculated and outlier statistics is used to identify differential binders. (D) The binding affinity of a protein to the RNA of interest can be determined by a dilution series of the RNA bait concentration. The RNA–bait concentration where half of the protein fraction is bound is referred to as the dissociation constant (K<sub>d</sub>), a measure of binding affinity.

proteins binding to specific RNA sequences [24], RNA secondary structures [25,26], and RNA modifications [8] (Figure 2B).

The broad use of RNA affinity purifications is due to the ease of performing the experiment since no genetic manipulations are needed. Recent advances in experimental procedures and mass spectrometry technologies enable affinity purifications in relatively high-throughput format, thereby increasing identification accuracy while lowering experimental costs [27]. Furthermore, affinity purifications using immobilized nucleic acids are not limited to determine binding specificity (regardless of whether an interaction occurs) (Figure 2C), but can also be applied to determine **binding affinity** (i.e., how strong an interaction is) (Figure 2D) [28].

Although affinity purifications are in principle easy to perform, care should be taken when designing an experiment to avoid false positive hits. Binding of proteins to RNA is influenced by various factors, such as sequence specificity and RNA secondary structures. It is possible that a particular RNA sequence adapts a different secondary structure *in vitro* compared with *in vivo* due to the immobilization linker sequence or the length of the RNA bait. Another problem is that the binding of some proteins depends on motifs that are located upstream or downstream of the motif of interest, which can be missed when using relatively short baits. Using longer RNA baits can circumvent these problems but, the longer the bait, the more proteins bind, which increases the background signal in the mass spectrometer. This can hinder the identification of low-abundant specific interactors by mass spectrometry. In addition, incubation and washing conditions can affect the identification of interactors, especially since affinity purifications do not capture weak and transient interactors well. Several experimental steps can be optimized to increase identification rates, such as preclearing of extracts [29,30] or using competitors during the experiment to eliminate nonspecific binding [22,24]. Furthermore, continuous improvements in the sensitivity and sequencing speed of modern mass spectrometers enable the identification of more proteins in shorter time frames, which increases the amount of information that can be obtained from single-step affinity purifications using crude lysates. Taken together, RNA affinity purifications are a powerful tool to study the molecular mechanisms of RNA–protein interactions, but experimental validations are essential to verify whether the detected *in vitro* interactions also occur *in vivo*.

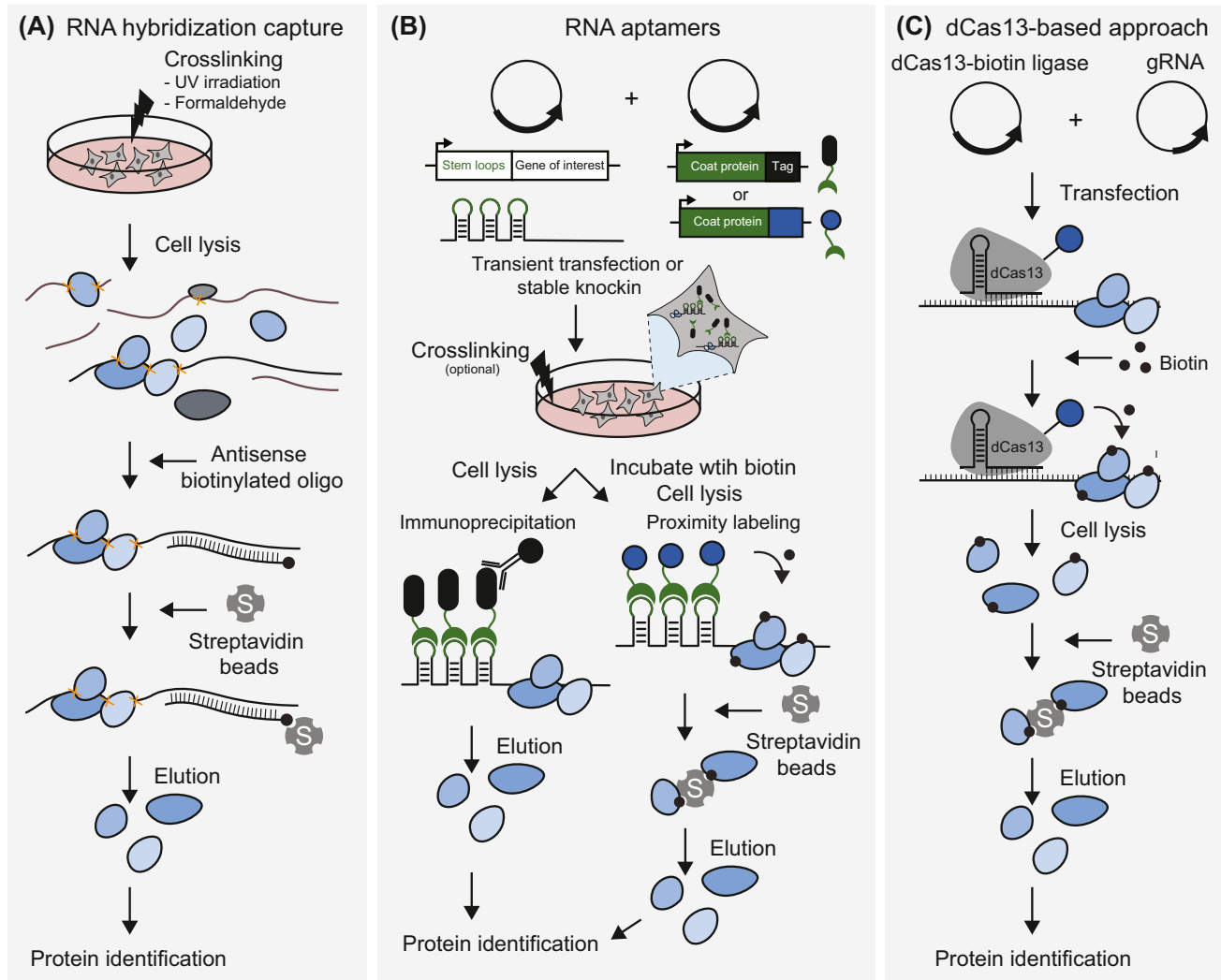
### **In Vivo Assembled RNA–Protein Complexes**

In addition to *in vitro*-based approaches, several methods have been developed to investigate *in vivo* assembled RNA–protein complexes. These methods can be divided into three categories: RNA hybridization; RNA-tags; or CRISPR-based RNA targeting (Figure 1D–F).

#### **RNA Hybridization Capture**

RNA antisense purification can be used to identify *trans*-genomic binding sites of endogenous RNAs [31–33], but in recent years this approach has also been used to identify RNA–protein interactions. In this method, RNA–protein interactions occurring in cells are first fixed by crosslinking, followed by capturing of an RNA of interest using specific biotin-tagged antisense DNA oligonucleotides. RNA–protein complexes are then purified using streptavidin-coated beads and, after extensive washes, bound proteins are eluted and identified by mass spectrometry (Figure 3A) [31,34–36].

In 2014, West and colleagues described a hybridization-capture approach, called capture hybridization analysis of RNA targets (CHART)-MS to study the interactome of the lncRNAs NEAT1 and MALAT1 [31]. One year later, three other hybridization-capture approaches, called comprehensive identification of RNA-binding proteins by mass spectrometry (ChIRP-MS) [34],



## Trends In Biotechnology

**Figure 3. Methods to Study *In Vivo* Assembled RNA-Protein Interactions.** (A) Schematic workflow of RNA hybridization capture. In brief, RNA-protein interactions in cells are preserved by crosslinking. The site of RNA-protein crosslinks are indicated in orange. After cell lysis, biotinylated DNA antisense oligonucleotides are hybridized to the target RNA, and the RNA-protein complexes are purified with streptavidin beads. The beads are washed and proteins are eluted. (B) RNA aptamer strategies rely on the co-expression of an aptamer-tagged RNA and coat proteins fused to a tag in living cells. The cells are crosslinked, lysed, and, subsequently, the RNA-associated proteins are captured by immunoprecipitation using the tag fused to the coat protein (left). Alternatively, the coat protein can be fused to a proximity-labeling enzyme, which, upon addition of biotin, biotinylates proteins in its proximity. The biotinylated proteins are isolated with streptavidin coated beads (right). (C) Workflow of catalytically dead Cas 13 (dCas13)-based methodology. Cells are transfected with dCas13 and guide RNAs (gRNAs). The gRNA recruits dCas13 fused to a proximity-labeling enzyme to an RNA of interest. Upon addition of biotin, the proximity-labeling enzyme biotinylates proteins in its proximity. These proteins are isolated with streptavidin beads and identified by mass spectrometry.

RNA antisense purification (RAP)-MS [35], and identification of direct RNA-interacting proteins (iDRiP) [36], were developed. Interestingly, all three studies applied their approaches to identify the interactome of Xist, a lncRNA involved in inactivation of the X chromosome in female cells. Strikingly, the overlap of identified Xist interacting proteins between these three methods is rather low, which might be due to methodological differences. These include use of different crosslinking reagents, the number and size of antisense oligonucleotides, and the use of different cutoffs to define statistically significant interactors (reviewed in [37]). This illustrates that the experimental

design can significantly affect which RNA–protein interactions are identified. Antisense oligonucleotides in particular need to be designed carefully so that they are specific for the target RNA and their binding site is accessible. However, once antisense oligonucleotides are designed and tested, a target RNA can be isolated from cells without genetic manipulations, which enables the identification of *in vivo* RNA–protein interactions also from primary cells and tissue samples. Hybridization capture approaches have been applied to a range of RNAs, successfully identifying RNA–protein interactions [38,39].

### RNA Tagging

The general strategy of RNA-tag-mediated purifications involves the attachment of stem loop structures, called **aptamers**, to the RNA of interest (Figure 3B). These introduced stem loops, often 4–24 repeats, are recognized by their corresponding RBP/**coat protein**. The most widely used aptamer is the MS2 stem loop originating from the bacteriophage MS2, which interacts with high specificity and affinity to the MS2 phage coat protein (MS2-CP) [40]. In mammalian cells, MS2 sequences do not naturally occur and, therefore, MS2-CPs do not recognize mammalian RNAs. Other aptamer–protein combinations exist, such as PP7 stem loop–PP7 coat protein [41] and BoxB stem loop– $\lambda$ N peptide [42]. The aptamer-tagged RNA recruits the coat proteins fused to peptide tags to facilitate the isolation of associated proteins. The method RBP purification and identification (RaPID) uses the interaction between MS2 and MS2-CP, which is fused to GFP and the streptavidin-binding peptide (SBP) tag [43]. The fluorescent label allows imaging of RNA and the SBP tag enables the purification of the RNA-associated proteins [43]. Several similar methods, such as MS2 *in vivo* biotin tagged RNA affinity purification (MS2-BioTRAP) [44], MS2-tagged RNA affinity purification (MS2-TRAP) [45,46], and MS2-tagged RNA affinity purification and mass spectrometry (MTRAP-MS) [47], have been described.

When designing RNA-tag-based purification, several considerations need to be considered. The choice of a small tag with high affinity and specificity is favored to avoid interference with the folding and structure of the coat protein, and to apply stringent washes (e.g., using washing buffers containing high salt and detergent concentrations) to remove nonspecific binders [48]. Furthermore, the expression level of the coat protein relative to the RNA expression level and the number of stem loop repeats is critical to achieve the best signal-to-noise ratio. It is unclear how many stem loops are required to achieve an optimal recruitment of the coat protein. Reducing the number of stem loops or even a single stem loop might be beneficial to minimize structural interference and binding of endogenous RBPs. Few methods reported the tagging of an RNA molecule with just a single stem loop [49,50]. Lee and colleagues used a single hairpin sequence that can be recognized by the CRISPR endonuclease Csy4 to study the interactome of three pre-miRNAs in different cell types [50]. This shows that mass spectrometry analysis of single hairpin-tagged RNAs is feasible.

Recently a different strategy, called in-cell protein–RNA interaction (incPRINT) was developed, which does not require mass spectrometry-based analysis [51]. The method relies on the co-expression of three components; recombinant luciferase tethered to MS2-CP; an MS2-tagged RNA; and a FLAG-tagged protein. The cell lysates are subjected to 384-well plates coated with anti-FLAG beads for immunoprecipitation and, together with the luminescence detection, the FLAG-tagged proteins that interact with the tested RNA can be identified. For systematic identification of RNA–protein interactions, a library of FLAG-tagged proteins is needed. As a proof of concept, three conserved regions of the murine Xist RNA were interrogated with ~3000 FLAG tagged proteins, identifying region-specific interactions for the ~17 kb-long Xist transcript [51].



### Proximity Proteome of Specific RNA Transcripts: Aptamer-Based Approaches

Further development of aptamer-based methods lies in the recruitment of **proximity-labeling enzymes** to the stem loop-tagged RNAs by fusing coat proteins to those enzymes. Proximity-labeling enzymes biotinylate accessible lysines or tyrosines of proteins within a 10–20-nm radius. The most commonly used enzymes are APEX and BioID (a promiscuous mutant of the BirA enzyme). The advantages of proximity-dependent biotinylation include the capture of weak and transient interactions as well as insoluble proteins because the biotinylation happens in the living cell (reviewed in [52]). Furthermore, due to the strong affinity between biotin and streptavidin-conjugated beads, stringent washes can be used to reduce the background signal. Besides labeling of proteins, APEX2 can also label RNA [53], enabling the parallel identification of RNAs and proteins that interact with the RNA of interest. In addition, biotin-labeling enzymes with a faster labeling time, such as APEX2 and (mini)TurboID, enable the study of dynamic changes in RNA–protein interactions over time [54]. Yet, proximity-dependent biotinylation also has some limitations, which are discussed later (see ‘Proximity Proteome of Specific RNA Transcripts: Future Directions’).

Ramanathan and colleagues developed the method RNA–protein interaction detection (RaPID), which involves a BoxB stem loop-tagged RNA and the  $\lambda$ N peptide fused to the biotin ligases BirA\* or BASU [55]. RaPID-western was applied to validate a known RNA–protein interaction between UDEN15 UG-rich RNA sequence and CELF1. Even though the methods cannot be directly compared, RaPID-western showed significant enrichment of CELF1 binding to UDEN15 compared with conventional RNA affinity purifications [55]. Others fused the MS2-CP to APEX2 to identify proteins interacting with the human telomerase RNA [56]. These studies investigated the interactome of overexpressed tagged RNAs rather than of endogenous RNA. However, recently RNA-BioID was applied to study the interactome of endogenously tagged  $\beta$ -actin RNA [57]. The paucity of studies focusing on endogenous interactions might have two reasons; first, endogenous tagging is challenging and, second, the RNA abundance of most transcripts is too low to provide sufficient enrichment.

### Proximity Proteome of Specific RNA Transcripts: CRISPR-Based Approaches

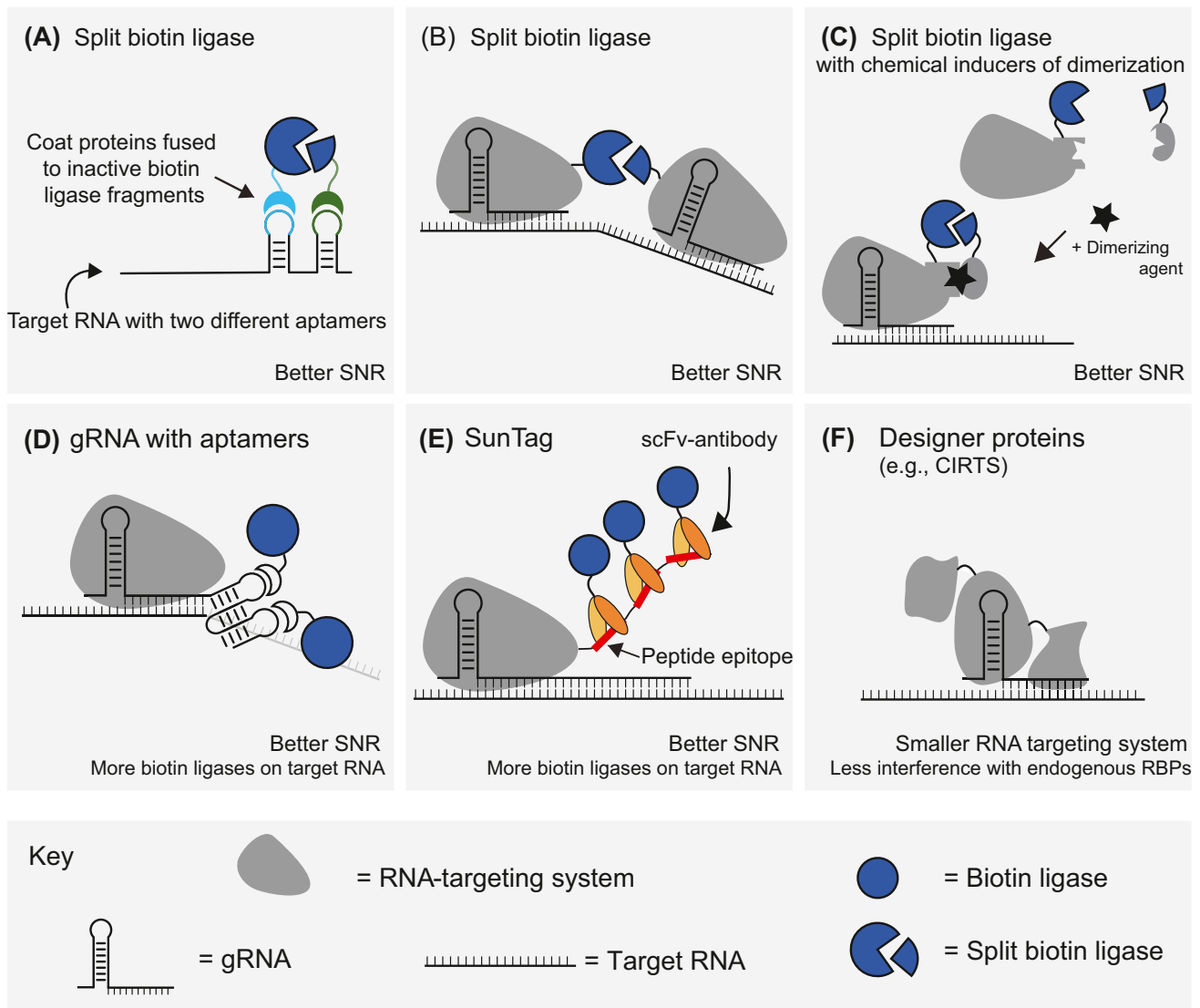
The discovery of the CRISPR-Cas13 RNA targeting system (Box 1) provides new opportunities to recruit proximity-labeling enzymes to an RNA transcript of interest. This new methodology has been recently applied by several groups [56,58–61]. Zhang and colleagues developed a

#### Box 1. Cas13: An Expansion of the CRISPR/Cas-Based Toolbox

CRISPR and CRISPR-associated (Cas) enzymes are part of the adaptive immune system in prokaryotes and Archaea, protecting them from foreign nucleic acids [78]. The CRISPR/Cas system became famous for its application in eukaryotic gene editing, where the precise and programmable DNA cleavage activity of Cas9 and Cas12a is used to introduce changes in the genome [79]. Since the first discovery of the CRISPR/Cas system, researchers have been studying and identifying new Cas enzymes to improve and expand their applications in biotechnology. For instance, recent computational approaches discovered class 2 type VI CRISPR/Cas enzymes that target RNA [80]. These enzymes belong to the Cas13 family, with at least six identified subtypes, Cas13a (previously known as C2C2), Cas13b, Cas13c, Cas13d, Cas13e, and Cas13f [81–83]. All identified Cas13 enzymes have two distinct ribonuclease activities. The first is important for processing pre-CRISPR RNAs generated by transcription of the CRISPR array into functional CRISPR RNAs that are complementary to the foreign RNA. Second, upon gRNA-dependent binding of Cas13 to the target RNA, the two higher eukaryotes and prokaryotes nucleotide-binding (HEPN) domains are activated and cleave the target RNA [84]. However, this second endonuclease activity is not only restricted to the target RNA, but also contributes to global RNA degradation. This phenomenon, called collateral cleavage, occurs in prokaryotic cells and *in vitro*, but further research is needed to understand the underlying molecular mechanisms [85,86]. Given that collateral cleavage is uncommon in eukaryotic cells and Cas13 shows precise targeting of RNA in these cells, Cas13 is a powerful tool for specific RNA cleavage with little to no off-target events [87,88]. The development of catalytically dead versions of Cas13 extended the applications beyond RNA knockdowns, allowing the recruitment of effector proteins to a RNA of interest, enabling imaging [89], editing [88], and splicing alterations [90] of specific RNAs.



CRISPR-based RNA-united interacting system (CRUIS), which is based on the fusion of **catalytically dead Cas 13a (dCas13a)** to the proximity-labeling enzyme PUP-IT. CRUIS was applied to identify the interactome of endogenous lncRNA NORAD and p21 mRNA [58]. By contrast, Yi and colleagues developed CRISPR-assisted RNA–protein interaction detection (CARPID) to detect protein interactors of Xist, DANCR, and MALAT1 by recruiting dCas13d-BASU [59]. Although these studies are based on the same principle, key differences are apparent,



Trends in Biotechnology

**Figure 4. Potential Refinements of Proximity Labeling-Based RNA–Protein Identification.** The use of a split proximity-labeling enzyme (A–C) could allow a better signal-to-noise-ratio (SNR). (A) For instance, by incorporating two different stem loops within the RNA of interest, an intact proximity-labeling enzyme is assembled only when bound to the target RNA. (B) Another option could be the delivery of two guides in the cell that are recognized by different hairpin-binding proteins [e.g., catalytically dead Cas 13 (dCas13b and dCas13d)] each fused to one half of the proximity-labeling enzyme, where the guides are spaced at certain distances to allow dimerization. (C) Alternatively, the ligase activity could be reconstituted by combining split-protein strategies with chemical inducers of dimerization. The recruitment of more biotin ligase to the target RNA can also improve the SNR (D,E). This can be achieved by (D) expressing a guide RNA (gRNA) containing aptamers that can recruit coat proteins fused to a proximity-labeling enzyme. (E) In the SunTag system, dCas13 can be fused to a peptide array, which can recruit multiple scFv-proximity-labeling enzymes. (F) Engineering smaller proteins that can target and bind to the RNA of interest might lead to decreased interference with other RNA-binding proteins (RBPs).

such as targeting endogenous or exogenous RNAs, expression levels of the target RNA, and the combination of dCas13 subtypes with proximity-labeling enzymes. Although the first studies using the CRISPR-Cas13 RNA target system to identify proteins interacting with specific RNA transcripts are promising, more insights into the underlying molecular mechanisms of this system are needed. For example, the design of effective guide RNAs (gRNAs) remains challenging because target recognition requirement is not yet fully understood and only a few tools exist to predict gRNA efficiency [62,63].

#### Proximity Proteome of Specific RNA Transcripts: Future Directions

A major challenge of proximity-dependent biotinylation methods, either aptamer or Cas13 based, is to improve the signal-to-noise ratio. Proximity-labeling enzymes facilitate biotinylation of proximal proteins even when not bound to the target RNA, resulting in biotinylation of nonspecific proteins. To overcome this limitation, the reconstitution of the proximity-labeling enzyme activity on the target RNA molecule might be a promising strategy [64]. Each half of the proximity-labeling enzyme can be fused to a different coat protein that reassembles when bound to their cognate stem loops inserted into the target RNA (Figure 4A). Another approach would be to use two gRNAs that are recognized by different hairpin-binding proteins each fused to one half of the proximity-labeling enzyme, where the gRNAs are spaced at a certain distance to allow dimerization (Figure 4B). Another way to reconstitute proximity-labeling activity on the target RNA could be achieved by combining split-protein strategies with chemical inducers of dimerization [65] (Figure 4C). Split enzymes have already been developed for APEX2, TurboID, and BioID, making the adaption of current methods to study RNA-protein interactions with split enzymes feasible [64,66–68].

Furthermore, increasing the number of proximity-labeling enzyme on the target RNA could improve the signal-to-noise ratio. As demonstrated for Cas9, designing gRNAs with additional aptamers results in recruitment of multiple effector proteins to the targeted site, thereby improving gene-editing efficiency [69,70]. Recently, Zhao and colleagues incorporated two MS2 stem loops into the Cas13 gRNA to recruit a methyltransferase, enabling efficient m6A editing [71] (Figure 4D). Another suggestion is to use the SunTag technology to recruit multiple proximity-labeling enzymes to a single dCas13 protein (Figure 4E). SunTag has been combined with dCas9 [72], although not yet with dCas13, and was shown to improve imaging of genomic loci, as well as gene activation and repression [72,73]. Taken together, these and other adaptations could improve the signal-to-noise ratio.

Cas13 proteins are rather large proteins (~100–140 kDa), which might influence the binding of endogenous RBPs to the Cas13-targeted RNA molecule. Therefore, smaller RNA-targeting systems would be beneficial (Figure 4F), such as the CRISPR/Cas-inspired RNA targeting system (CIRTS, ~25 kDa) [74]. The CIRTS system is, similar to Cas13, a gRNA-dependent RNA targeting system that could deliver proximity-labeling enzymes to an RNA of interest. Further technical innovations will facilitate the development of robust methods to comprehensively identify interactomes for specific RNA transcripts.

#### Concluding Remarks

Taken together, a variety of RNA-centric methods exist to study the proteome of a given RNA transcript, yielding new insights into RNA biology. Recent developments of dCas13-based approaches will facilitate the identification of interactomes for specific RNA transcript, possibly even in a high-throughput manner. This could be applied, for example, to map proteins binding across the full length of an RNA transcript of interest (see Outstanding Questions). Furthermore, combining RNA-centric methods with complementary approaches is required to validate

#### Outstanding Questions

How can we capture the complete protein interactome of a specific RNA transcript?

Is it possible to map the proteins along an RNA transcript of interest and, if so, how?

What is the overlap of identified RBPs when using different RNA-centric methods studying the same RNA transcript?

Are dCas13-biotinylation-based approaches the future for identifying RNA-protein interactions, or are there other novel targeting or labeling systems?

How can we apply the knowledge of RNA-protein interactions to various fields, such as therapeutics and bioengineering?

identified interactors and to gain a deeper understanding of RNA biology. Even more insights into the complexity of RNA interactions can be gained by studying RNA–RNA, RNA–DNA, or RNA–chromatin interactions.

Given that RNA molecules have a central role in disease, RNA–protein interactions are considered an attractive target for therapy. For example, small molecules have been developed to target the interaction between transactivator of transcription (TAT) protein and transactivator response RNA (TAR) in HIV, reducing viral replication [75,76]. Furthermore, knowledge of RNA–protein interactions can be used to design proteins for various purposes, including RNA degradation, editing, or imaging [77]. Thus, further development and implementation of technologies to identify RNA–protein interactions will have a profound impact, not only from a fundamental, but also from an applied scientific perspective.

### Acknowledgments

We apologize to authors whose work could not be cited due to space constraints. C.G., S.S., F.A.H.v.H., and M.V. are part of the Oncode Institute, which is partly funded by the Dutch Cancer Society (KWF).

### References

- Morris, K.V. and Mattick, J.S. (2014) The rise of regulatory RNA. *Nat. Rev. Genet.* 15, 423–437
- Strobel, E.J. *et al.* (2016) RNA systems biology: uniting functional discoveries and structural tools to understand global roles of RNAs. *Curr. Opin. Biotechnol.* 39, 182–191
- Steffl, R. *et al.* (2005) RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. *EMBO Rep.* 6, 33–38
- Beckmann, B.M. *et al.* (2016) The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. *Pflügers Arch. Eur. J. Physiol.* 468, 1029–1040
- Hentze, M.W. *et al.* (2018) A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* 19, 327–341
- Lukong, K.E. *et al.* (2008) RNA-binding proteins in human genetic disease. *Trends Genet.* 24, 416–425
- Castello, A. *et al.* (2013) RNA-binding proteins in Mendelian disease. *Trends Genet.* 29, 318–327
- Edupuganti, R.R. *et al.* (2017) N6-methyladenosine (m6A) recruits and repels proteins to regulate mRNA homeostasis. *Nat. Struct. Mol. Biol.* 24, 870–878
- Weng, H. *et al.* (2018) METTL14 inhibits hematopoietic stem/progenitor differentiation and promotes leukemogenesis via mRNA m6A modification. *Cell Stem Cell* 22, 191–205.e9
- Elcheva, I.A. and Spiegelman, V.S. (2020) Targeting RNA-binding proteins in acute and chronic leukemia. *Leukemia* Published online November 4, 2020. <https://doi.org/10.1038/s41375-020-01066-4>
- Ule, J. *et al.* (2003) CLIP Identifies Nova-regulated RNA networks in the brain. *Science* 302, 1212–1215
- Van Nostrand, E.L. *et al.* (2020) A large-scale binding and functional map of human RNA-binding proteins. *Nature* 583, 711–719
- Shchepachev, V. *et al.* (2019) Defining the RNA interactome by total RNA-associated protein purification. *Mol. Syst. Biol.* 15, e8689
- Queiroz, R.M.L. *et al.* (2019) Comprehensive identification of RNA–protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nat. Biotechnol.* 37, 169–178
- Trendel, J. *et al.* (2019) The human RNA-binding proteome and its dynamics during translational arrest. *Cell* 176, 391–403.e19
- Urdaneta, E.C. *et al.* (2019) Purification of cross-linked RNA–protein complexes by phenol-toluol extraction. *Nat. Commun.* 10, 990
- Castello, A. *et al.* (2012) Insights into RNA biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* 149, 1393–1406
- Baltz, A.G. *et al.* (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* 46, 674–690
- Liepert, A. *et al.* (2016) Identification of RNA-binding proteins in macrophages by interactome capture. *Mol. Cell. Proteomics* 15, 2699–2714
- Perez-Perri, J.I. *et al.* (2018) Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. *Nat. Commun.* 9, 4408
- Castello, A. *et al.* (2016) Comprehensive identification of RNA-binding domains in human cells. *Mol. Cell* 63, 696–710
- Tacheny, A. *et al.* (2013) Mass spectrometry-based identification of proteins interacting with nucleic acids. *J. Proteome* 94, 89–109
- Grabowski, P.J. and Sharp, P.A. (1986) Affinity chromatography of splicing complexes: U2, U5, and U4+U6 small nuclear ribonucleoprotein particles in the spliceosome. *Science* 233, 1294–1299
- Butter, F. *et al.* (2009) Unbiased RNA-protein interaction screen by quantitative proteomics. *Proc. Natl. Acad. Sci. U. S. A.* 106, 10626–10631
- Serikawa, T. *et al.* (2018) Comprehensive identification of proteins binding to RNA G-quadruplex motifs in the 5' UTR of tumor-associated mRNAs. *Biochimie* 144, 169–184
- Herdry, B. *et al.* (2018) Analysis of NRAS RNA G-quadruplex binding proteins reveals DDX3X as a novel interactor of cellular G-quadruplex containing transcripts. *Nucleic Acids Res.* 46, 11592–11604
- Hubner, N.C. *et al.* (2015) A quantitative proteomics tool to identify DNA-protein interactions in primary cells or blood. *J. Proteome Res.* 14, 1315–1329
- Makowski, M.M. *et al.* (2018) Global profiling of protein-DNA and protein-nucleosome binding affinities using quantitative mass spectrometry. *Nat. Commun.* 9, 1653
- Ward, A.M. *et al.* (2011) Quantitative mass spectrometry of DENV-2 RNA-interacting proteins reveals that the DEAD-box RNA helicase DDX6 binds the DB1 and DB2 3' UTR structures. *RNA Biol.* 8, 984–987
- Iioka, H. *et al.* (2011) Efficient detection of RNA-protein interactions using tethered RNAs. *Nucleic Acids Res.* 39, e53
- West, J.A. *et al.* (2014) The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell* 55, 791–802
- Chu, C. *et al.* (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA–chromatin interactions. *Mol. Cell* 44, 667–678
- Engreitz, J.M. *et al.* (2013) The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 341, 1237973
- Chu, C. *et al.* (2015) Systematic discovery of Xist RNA binding proteins. *Cell* 161, 404–416
- McHugh, C.A. *et al.* (2015) The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* 521, 232–236
- Minajigi, A. *et al.* (2015) A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* 349, aab2276

37. Da Rocha, S.T. and Heard, E. (2017) Novel players in X inactivation: insights into Xist-mediated gene silencing and chromosome conformation. *Nat. Struct. Mol. Biol.* 24, 197–204
38. Knoener, R.A. *et al.* (2017) Elucidating the in vivo interactome of HIV-1 RNA by hybridization capture and mass spectrometry. *Sci. Rep.* 7, 16965
39. Fanucchi, S. *et al.* (2019) Immune genes are primed for robust transcription by proximal long noncoding RNAs located in nuclear compartments. *Nat. Genet.* 51, 138–150
40. Johansson, H.E. *et al.* (1998) A thermodynamic analysis of the sequence-specific binding of RNA by bacteriophage MS2 coat protein. *Proc. Natl. Acad. Sci. U. S. A.* 95, 9244–9249
41. Lim, F. and Peabody, D.S. (2002) RNA recognition site of PP7 coat protein. *Nucleic Acids Res.* 30, 4138–4144
42. Austin, R.J. *et al.* (2002) Designed arginine-rich RNA-binding peptides with picomolar affinity. *J. Am. Chem. Soc.* 124, 10966–10967
43. Slobodin, B. and Gerst, J.E. (2011) RaPID: an aptamer-based mRNA affinity purification technique for the identification of RNA and protein factors present in ribonucleoprotein complexes. *Methods Mol. Biol.* 714, 387–406
44. Tsai, B.P. *et al.* (2011) Quantitative profiling of in vivo-assembled RNA-protein complexes using a novel integrated proteomic approach. *Mol. Cell. Proteomics* 10, M110.007385
45. Yoon, J.H. *et al.* (2012) MS2-TRAP (MS2-tagged RNA affinity purification): Tagging RNA to identify associated miRNAs. *Methods* 58, 81–87
46. Yoon, J.H. and Gorospe, M. (2016) Identification of mRNA-interacting factors by MS2-TRAP (MS2-tagged RNA affinity purification). *Methods Mol. Biol.* 1421, 15–22
47. Liu, S. *et al.* (2015) Identification of lncRNA MEG3 binding protein using MS2-tagged RNA affinity purification and mass spectrometry. *Appl. Biochem. Biotechnol.* 176, 1834–1845
48. Gemmill, D. *et al.* (2020) Current approaches for RNA-labelling to identify RNA-binding proteins. *Biochem. Cell Biol.* 98, 31–41
49. Cottrell, K.A. and Djuranovic, S. (2016) Urb-RIP - an adaptable and efficient approach for immunoprecipitation of RNAs and associated RNAs/proteins. *PLoS ONE* 11, e0167877
50. Lee, H.Y. *et al.* (2013) RNA-protein analysis using a conditional CRISPR nuclease. *Proc. Natl. Acad. Sci. U. S. A.* 110, 5416–5421
51. Graindorge, A. *et al.* (2019) In-cell identification and measurement of RNA-protein interactions. *Nat. Commun.* 10, 5317
52. Samavarchi-Tehrani, P. *et al.* (2020) Proximity dependent biotinylation: Key enzymes and adaptation to proteomics approaches. *Mol. Cell. Proteomics* 19, 757–773
53. Fazal, F.M. *et al.* (2019) Atlas of subcellular RNA localization revealed by APEX-Seq. *Cell* 178, 473–490.e26
54. Cho, K.F. *et al.* (2020) Proximity labeling in mammalian cells with TurboID and split-TurboID. *Nat. Protoc.* 15, 3971–3999
55. Ramanathan, M. *et al.* (2018) RN A-protein interaction detection in living cells. *Nat. Methods* 15, 207–212
56. Han, S. *et al.* (2020) RNA-protein interaction mapping via MS2 or Cas13-based APEX targeting. *Proc. Natl. Acad. Sci. U. S. A.* 117, 22068–22079
57. Mukherjee, J. *et al.* (2019)  $\beta$ -Actin mRNA interactome mapping by proximity biotinylation. *Proc. Natl. Acad. Sci. U. S. A.* 116, 12863–12872
58. Zhang, Z. *et al.* (2020) Capturing RNA-protein interaction via CRUIS. *Nucleic Acids Res.* 48, e52
59. Yi, W. *et al.* (2020) CRISPR-assisted detection of RNA-protein interactions in living cells. *Nat. Methods* 17, 685–688
60. Lin, X. and Lawrenson, K. (2020) In vivo analysis of RNA proximity proteomes using RiboPro. *bioRxiv* Published online February 28, 2020. <https://doi.org/10.1101/2020.02.28.970442>
61. Li, Y. *et al.* (2020) CBRPP: a new RNA-centric method to study RNA-protein interactions. *bioRxiv* Published online April 9, 2020. <https://doi.org/10.1101/2020.04.09.033290>
62. Labun, K. *et al.* (2019) CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* 47, W171–W174
63. Wessels, H.H. *et al.* (2020) Massively parallel Cas13 screens reveal principles for guide RNA design. *Nat. Biotechnol.* 38, 722–727
64. Han, Y. *et al.* (2019) Directed evolution of split APEX2 peroxidase. *ACS Chem. Biol.* 14, 619–635
65. Shekhawat, S.S. and Ghosh, I. (2011) Split-protein systems: beyond binary protein-protein interactions. *Curr. Opin. Chem. Biol.* 15, 789–797
66. Cho, K.F. *et al.* (2020) Split-TurboID enables contact-dependent proximity labeling in cells. *Proc. Natl. Acad. Sci. U. S. A.* 117, 12143–12154
67. Schopp, I.M. *et al.* (2017) Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes. *Nat. Commun.* 8, 15690
68. De Munter, S. *et al.* (2017) Split-BioID: a proximity biotinylation assay for dimerization-dependent protein interactions. *FEBS Lett.* 591, 415–424
69. Konermann, S. *et al.* (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517, 583–588
70. Zalatan, J.G. *et al.* (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* 160, 339–350
71. Zhao, J. *et al.* (2020) Photoactivatable RNA N6-methyladenosine editing with CRISPR-Cas13. *Small* 16, 1907301
72. Tanenbaum, M.E. *et al.* (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159, 635–646
73. Chavez, A. *et al.* (2016) Comparison of Cas9 activators in multiple species. *Nat. Methods* 13, 563–567
74. Rauch, S. *et al.* (2019) Programmable RNA-guided RNA effector proteins built from human parts. *Cell* 178, 122–134
75. Mei, H.Y. *et al.* (1997) Discovery of selective, small-molecule inhibitors of RNA complexes - I. The Tat protein/TAR RNA complexes required for HIV-1 transcription. *Bioorganic Med. Chem.* 5, 1173–1184
76. Mei, H.Y. *et al.* (1998) Inhibitors of protein-RNA complexation that target the RNA: Specific recognition of human immunodeficiency virus type 1 TAR RNA by small organic molecules. *Biochemistry* 37, 14204–14212
77. Zhao, Y.Y. *et al.* (2018) Expanding RNA binding specificity and affinity of engineered PUF domains. *Nucleic Acids Res.* 46, 4771–4782
78. Makarova, K.S. *et al.* (2011) Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* 9, 467–477
79. Cong, L. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823
80. Shmakov, S. *et al.* (2015) Discovery and functional characterization of diverse Class 2 CRISPR-Cas systems. *Mol. Cell* 60, 385–397
81. Shmakov, S. *et al.* (2017) Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.* 15, 169–182
82. Yan, W.X. *et al.* (2018) Cas13d is a compact RNA-targeting type VI CRISPR effector positively modulated by a WYL-domain-containing accessory protein. *Mol. Cell* 70, 327–339
83. Xu, C. *et al.* (2020) Novel miniature CRISPR-Cas13 systems from uncultivated microbes effective in degrading SARS-CoV-2 sequences and influenza viruses. *Res. Sq.* Published online May 21, 2020. <http://dx.doi.org/10.21203/rs.3.rs-30924/v1>
84. O'Connell, M.R. (2019) Molecular mechanisms of RNA targeting by Cas13-containing Type VI CRISPR-Cas systems. *J. Mol. Biol.* 431, 66–87
85. East-Seletsky, A. *et al.* (2016) Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* 538, 270–273
86. Abudayyeh, O.O. *et al.* (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353, aaf5573
87. Abudayyeh, O.O. *et al.* (2017) RNA targeting with CRISPR-Cas13. *Nature* 550, 280–284
88. Cox, D.B.T. *et al.* (2017) RNA editing with CRISPR-Cas13. *Science* 358, 1019–1027
89. Yang, L.Z. *et al.* (2019) Dynamic imaging of RNA in living cells by CRISPR-Cas13 systems. *Mol. Cell* 76, 981–997
90. Konermann, S. *et al.* (2018) Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. *Cell* 173, 665–676